

# Doubly Robust-Based Generalized Estimating Equations for the Analysis of Longitudinal Ordinal Missing Data

José Luiz P. da Silva, Enrico A. Colosimo, Fábio N. Demarqui

Departament of Statistics

Universidade Federal de Minas Gerais

June 16, 2015

## Abstract

Generalized Estimation Equations (GEE) are a well-known method for the analysis of non-Gaussian longitudinal data. This method has computational simplicity and marginal parameter interpretation. However, in the presence of missing data, it is only valid under the strong assumption of missing completely at random (MCAR). Some corrections can be done when the missing data mechanism is missing at random (MAR): inverse probability weighting (WGEE) and multiple imputation (MIGEE). In order to obtain consistent estimates, it is necessary the correct specification of the weight model for WGEE or the imputation model for the MIGEE. A recent method combining ideas of these two approaches has doubly robust property. For consistency, it requires only the weight or the imputation model to be correct. In this work it is assumed a proportional odds model and it is proposed a doubly robust estimator for the analysis of ordinal longitudinal data with intermittently missing response and covariate under the MAR mechanism. Simulation results revealed better performance of the proposed method compared to WGEE and MIGEE. The method is applied to a data set related to Analgesia Pain in Childbirth study.

Keywords: Missing at random; Multiple imputation; Proportional Odds Model; Weighted GEE.

## 1 Introduction

The Generalized Estimating Equation (GEE) method (Liang & Zeger, 1986) is one of the most popular approaches for the analysis of non-Gaussian correlated data. Its main advantage resides in the fact that one is only required to specify correctly the mean structure of the response for the parameter estimator to be consistent and asymptotically normal. In its basic formulation the association parameters among repeated measures were taken as nuisance parameters. GEE method has computational simplicity (there is no need of dealing with complex, and in some cases, intractable likelihoods) and it further allows populational-averaged interpretation of the parameter of interest.

It is very common for sets of longitudinal data to be incomplete, in the sense that not all planned observations are actually observed. This problem is pervasive in longitudinal data because nonresponse can occur any time from the beginning of the study. Two patterns of missing data can be observed for the response: (1) dropout, when a subject leaves the study prematurely for reasons beyond the control of the investigator, leading to a monotone pattern of nonresponse, or (2) intermittent nonresponse, in which a subject returns to the study after some occasions of nonresponse. Covariates may also be missing, leading to limitations in data analysis. In the presence of missing data, three issues are of main concern: (1) potential serious bias due to systematic differences between the observed data and the missing data, (2) complications in data handling and statistical inferences, and (3) loss of efficiency. Therefore, in order to make valid inferences it is fundamental to know the missing data mechanism generating the nonresponse and how to handle it.

Little & Rubin (1987) provided a formal framework for dealing with missing data by defining the commonly adopted taxonomy of missing data mechanisms. A nonresponse process is said to be *missing completely at random* (MCAR) if missingness is independent of both unobserved and observed data, and *missing at random* (MAR) if, conditionally on the observed data, the missingness is independent of the unobserved data. When the nonresponse process depends on unobserved quantities it is said to be *missing not at random* (MNAR).

When data are incomplete, GEE suffers from its frequentist nature and is, in its basic form, valid only under MCAR (Liang & Zeger, 1986). The first effort to make GEE applicable to the more realistic MAR scenario was Multiple Imputation (MIGEE), proposed by Little & Rubin (1987), in which the missing portions of data are multiply imputed taking into account the uncertainty associated with the predicted values. The completed data sets are analyzed by standard methods for complete data, and estimates are combined into a final analysis. Multiple imputation is detailed in the books by Schafer (1997), Little & Rubin (2002) and Carpenter & Kenward (2013). Later, Robins *et al.* (1995) proposed the Weighted Generalized Estimating Equations (WGEE), which consists in weighting each observation by the inverse of the probability of the data being observed. This method produces consistent estimates provided the weight model is correctly specified.

Doubly robust estimators (DRGEE) arise as a third generalization of ordinary GEE to deal with data subject to MAR mechanism. DR methods have received increasingly attention in the literature in the last decade (see Carpenter *et al.* (2006), Bang & Robins (2005), Tsiatis (2006), Seaman & Copas (2009), Chen & Zhou (2011)). The main idea is to supplement the WGEE with a predictive model for the missing quantities conditional on the observed ones. Doubly Robust method requires only the dropout or the conditional model to be correctly specified in order to provide consistent estimates. In the analysis of longitudinal binary data, DR methods have been applied by Seaman & Copas (2009), Birhanu *et al.* (2011), for missing responses and by Chen & Zhou (2011) for intermittently missing response and a single missing covariate.

Literature of GEE for missing data is comparatively scarce for longitudinal ordinal response. In Toledano & Gatsonis (1999), the authors used a weighted GEE method to accommodate arbitrary patterns of a MCAR missing response and missingness in a key covariate subject to a MAR mechanism. A recent paper from Donneau *et al.* (2014a) compared through a simulation study two multiple imputation methods (multivariate normal

imputation and ordinal imputation regression) for longitudinal ordinal data subject to dropout. In another paper the same authors compared joint modeling and fully conditional specification approaches for non-monotone missingness (Donneau *et al.*, 2014b). The above mentioned papers used single robust versions of GEE and they have treated only a missing MAR baseline covariate or missing MAR response. Thus the use of a DR GEE method for ordinal data with simultaneously intermittently missing response and missing covariate has been in need of further development.

This work was motivated by the Analgesia in Childbirth study which was conducted in Minas Gerais state, Brazil. The main objective of that study was to compare two techniques of analgesia for labor pain in 49 patients. The response, pain intensity, was subjectively assessed by each patient, and various clinical covariates were taken until delivery. Response and a particular covariate (consumption of oxytocin) were missing for some patients and the MAR mechanism seems to be a reasonable assumption for this data.

In the current paper, it is proposed a doubly robust approach for the analysis of longitudinal ordinal data with intermittently missing response and covariate that is MAR. The proposed methodology, to our knowledge new to the GEE literature of missing ordinal data, can be used for handling arbitrary patterns of missing data in the ordinal response and missingness in a key covariate, as those frequently arising in medical studies.

The paper is organized as follows. In Section 2 are defined the notation for GEE with fully observed data and missing data mechanisms. Section 3 outlines WGEE and MIGEE approaches. The proposed methodology is established in Section 4. A simulation study is presented in Section 5, in which the finite-sample biases and standard errors are compared for the standard GEE, MIGEE, WGEE and doubly robust versions. Data arising from the Analgesia en Childbirth study are analyzed in Section 6. Paper ends with a discussion and future directions in Section 7.

## 2 GEE for Complete Data and Missing Data Assumptions

In this section it is introduced generalized estimating equations for the analysis of fully observed ordinal data. Section 2.1 establishes the model and notation for longitudinal ordinal data. Section 2.2 presents a series of assumptions related to mechanism causing data to be missing and necessary to be considered in order to build valid estimators.

### 2.1 GEE for Longitudinal Ordinal Response

Let  $O_{it} \in \{1, 2, \dots, J\}$  be the ordinal response for subject  $i$  ( $i = 1, \dots, n$ ) at time  $t$  ( $t = 1, \dots, T_i$ ,  $T_i \leq T$ ). As the response has  $J$  levels it can be defined  $Y_{itj} = I(O_{it} = j)$  for  $j = 1, \dots, J$ , where  $I(A)$  denotes the indicator function.  $Y_{itj}$  is converted into the equivalent  $(J - 1)$ -variate vector  $\mathbf{Y}_{it} = (Y_{it1}, \dots, Y_{it(J-1)})^T$  and let  $\mathbf{Y}_i = (Y_{i1}^T, \dots, Y_{iT_i}^T)^T$  the stacked response vector. When  $J = 2$  the response is binary and  $\mathbf{Y}_{it}$  is a scalar. Let  $\mathbf{X}_i = (X_{i1}^T, \dots, X_{iT_i}^T)^T$  denotes the  $T_i \times 1$  covariate vector that may be missing for the  $i$ -th subject, and  $\mathbf{Z}_i = (\mathbf{Z}_{i1}^T, \dots, \mathbf{Z}_{iT_i}^T)^T$  the  $T_i \times q$  matrix of explanatory variables that are always observed.

The marginal distribution of  $\mathbf{Y}_{it}$  is assumed to be multinomial (with sample size  $\sum_{j=1}^J Y_{itj} = 1$ ), that is

$$f(\mathbf{Y}_{it}|\mathbf{X}_{it}, \mathbf{Z}_{it}, \boldsymbol{\beta}) = \prod_{j=1}^J \mu_{itj}^{Y_{itj}}, \quad (1)$$

where  $\mu_{itj} = \mu_{itj}(\boldsymbol{\beta}) = E(Y_{itj}|\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\beta}) = Pr(O_{it} = j|\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\beta})$ , is the probability of response  $j$  at time  $t$  and  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of parameters. Two common choices for modeling  $\mu_{itj}$  are the cumulative logit and probit models. In this work it is assumed a cumulative logit link, that is,

$$\text{logit}[Pr(O_{it} \leq j|X_{it}, Z_{it})] = \beta_{0j} + X_{it}\beta_x + \mathbf{Z}_{it}^T \boldsymbol{\beta}_z, \quad j = 1, \dots, J-1. \quad (2)$$

Formulation in (2) implies a proportional odds model (McCullagh, 1980). In such model the interpretation of  $\boldsymbol{\beta}$  is the same regardless of the number of categories (i.e., it is invariant to combination of categories). A desired feature is that the exponential of the parameters is interpreted as an odds ratio (Agresti, 2013).

Main interest is to make inferences related to the regression parameters  $\boldsymbol{\beta} = (\beta_{01}, \dots, \beta_{0,J-1}, \beta_x, \boldsymbol{\beta}_z^T)^T$  associated with the  $(J-1) \times 1$  marginal probability vectors

$$E(Y_{it}|\mathbf{X}_i, \mathbf{Z}_i) = \boldsymbol{\mu}_{it}(\boldsymbol{\beta}) = (\mu_{it1}, \dots, \mu_{it(J-1)})^T.$$

$\boldsymbol{\mu}_{it}$  is grouped to form a vector  $E(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i) = \boldsymbol{\mu}_i = (\boldsymbol{\mu}_{i1}^T, \dots, \boldsymbol{\mu}_{iT_i}^T)^T$  with the same dimension of  $\mathbf{Y}_i$ .

In order to estimate  $\boldsymbol{\beta}$  generalized estimation equations are used (Liang & Zeger (1986); Lipsitz *et al.* (1994)), which takes the form

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{U}_i(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \quad (3)$$

where  $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^T}$  and  $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$  is a  $T_i(J-1) \times T_i(J-1)$  “working covariance” matrix usually decomposed into the form  $\mathbf{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{F}_i^{1/2}(\boldsymbol{\beta}) \mathbf{C}_i(\boldsymbol{\alpha}) \mathbf{F}_i^{1/2}(\boldsymbol{\beta})$ , where  $\mathbf{F}_i$  is a matrix containing the marginal variances,  $\mathbf{F}_{it}$ , given by

$$\mathbf{F}_{it} = \text{diag}[\mu_{it1}(1 - \mu_{it1}), \dots, \mu_{it,J-1}(1 - \mu_{it,J-1})],$$

and  $\mathbf{C}_i$  is equal to the marginal correlation matrix. The  $(J-1) \times (J-1)$  diagonal blocks of  $\mathbf{C}_i(\boldsymbol{\alpha})$  are  $\mathbf{F}_{it}^{-1/2} \mathbf{V}_{it} \mathbf{F}_{it}^{-1/2}$ , with  $\mathbf{V}_{it} = \text{diag}(\boldsymbol{\mu}_{it}) - \boldsymbol{\mu}_{it} \boldsymbol{\mu}_{it}^T$ ; and the  $(J-1) \times (J-1)$  off-diagonal blocks of  $\mathbf{C}_i(\boldsymbol{\alpha})$  are  $\boldsymbol{\alpha}_{itt'}$ , which represents the correlation between  $\mathbf{Y}_{it}$  and  $\mathbf{Y}_{it'}$ ,  $t \neq t'$  (Lipsitz *et al.*, 1994).

Under mild regularity conditions and correct specification of the marginal mean model in (2), Liang and Zeger (1986) proved that the estimator  $\hat{\boldsymbol{\beta}}$ , obtained by solving (3), is consistent and  $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$  converges in distribution to a  $p$ -variate normal distribution with mean  $\mathbf{0}$  and covariance matrix

$$\mathbf{V}_{\boldsymbol{\beta}} = \lim_{n \rightarrow \infty} n \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_0^{-1}, \quad (4)$$

where  $\boldsymbol{\Sigma}_0 = \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{D}_i^T$ , and  $\boldsymbol{\Sigma}_1 = \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} \text{Cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i^T$ . In practice, the “sandwich” covariance matrix  $\mathbf{V}_{\boldsymbol{\beta}}$  in (4) is calculated by ignoring the limit and replacing  $(\boldsymbol{\beta}, \boldsymbol{\alpha})$  and  $\text{Cov}(\mathbf{Y}_i)$  by  $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$  and  $(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)^T$ , respectively (Touloumis *et al.*, 2013).

## 2.2 Missing Data Framework

For each occasion  $t$  it can be defined  $R_{it} = 0$  if  $O_{it}$  and  $X_{it}$  are missing,  $R_{it} = 1$  if  $O_{it}$  is missing and  $X_{it}$  is observed,  $R_{it} = 2$  if  $O_{it}$  is observed and  $X_{it}$  is missing, and  $R_{it} = 3$  if  $O_{it}$  and  $X_{it}$  are both observed. Let  $\mathbf{R}_i = (R_{i1}, \dots, R_{iT_i})^T$ , and  $\bar{\mathbf{R}}_{it} = (R_{i1}, \dots, R_{i,t-1})$ .

By specifying conditional models of the form  $Pr(R_{it} = r_{it} | \bar{\mathbf{R}}_{it}, \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i)$  it can be obtained  $Pr(\mathbf{R}_i = \mathbf{r}_i | \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i)$  through  $\prod_{i=2}^{T_i} Pr(R_{it} = r_{it} | \bar{\mathbf{R}}_{it}, \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i) Pr(R_{i1} = r_{i1} | \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i)$ . Let  $\lambda_{itk} = Pr(R_{it} = k | \bar{\mathbf{R}}_{it}, \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i)$ , for  $k = 0, 1, 2, 3$ . This general formulation encompasses MCAR, MAR and MNAR mechanisms. In particular, the MAR mechanism requires

$$Pr(\mathbf{R}_i = \mathbf{r}_i | \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i) = Pr(\mathbf{R}_i = \mathbf{r}_i | \mathbf{O}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i), \quad (5)$$

where  $\mathbf{O}_i^o$  and  $\mathbf{X}_i^o$  denotes the observed components of  $\mathbf{O}_i$  and  $\mathbf{X}_i$ , respectively. Because  $\mathbf{R}_i$  is modeled through a product of conditional models, it is natural to make the following further assumption (Chen & Zhou, 2011)

$$Pr(R_{it} = r_{it} | \bar{\mathbf{R}}_{it}, \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i) = Pr(R_{it} = r_{it} | \bar{\mathbf{R}}_{it}, \bar{\mathbf{O}}_{it}^o, \bar{\mathbf{X}}_{it}^o, \bar{\mathbf{Z}}_{it}), \quad (6)$$

for each time  $t$ , where  $\bar{\mathbf{O}}_{it}^o$  and  $\bar{\mathbf{X}}_{it}^o$  are the histories of observed responses and covariates up to time  $t - 1$ .

Let  $\pi_{it} = Pr(R_{it} = 3 | \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i)$  be the marginal probability of observing both  $\mathbf{O}_i$  and  $\mathbf{X}_i$  at time  $t$ , given the entire vectors of responses and covariates. Then,  $\pi_{it}$  is expressed by

$$\pi_{it} = \sum_{r_{i1}, \dots, r_{i,t-1}} Pr(R_{it} = 3, R_{i,t-1} = r_{i,t-1}, \dots, R_{i1} = r_{i1} | \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i).$$

This marginal probability can be expressed in terms of the conditional probabilities  $\lambda_{itk}$ 's. Throughout the paper it is required the so-called *positivity assumption*, that is,  $\pi_{it}$  must be bounded away from zero. This condition is needed in order to guarantee the existence of  $\sqrt{n}$ -consistent estimators of  $\beta$  (Robins *et al.*, 1995).

## 3 Available Approaches for Missing Data

Multiple imputation and weighted generalized estimation equations are two commonly used methods available for missing data under MAR mechanism. These methods are presented in Sections 3.1 and 3.2, respectively. They serve as the basis for the construction of the doubly robust estimator, presented in Section 4.

### 3.1 Multiple Imputation Generalized Estimating Equations

A imputation model commonly used to handle intermittently missing response and covariate, is imputation using chained equations (van Buuren *et al.* (1999), van Buuren (2007)), which is more commonly referred to as full conditional specification (FCS). This approach specifies conditional distributions for each incomplete variable, conditional on all others variables in the imputation model. Starting from an initial imputation, FCS draws imputations by iterating over the conditional densities.

Denote by  $\tilde{\boldsymbol{\beta}}_m$  and  $\tilde{\mathbf{U}}_m$ , respectively, the estimate of  $\boldsymbol{\beta}$  and its covariance matrix from the GEE analysis of the  $m$ -th completed data set, ( $m = 1, \dots, M$ ). Following Rubin (1987), the combined point estimate for the parameter of interest  $\boldsymbol{\beta}$  from the MI is simply the average of the  $M$  complete-data point estimates

$$\hat{\boldsymbol{\beta}}_{MI} = \frac{1}{M} \sum_{m=1}^M \tilde{\boldsymbol{\beta}}_m,$$

and an estimate of the covariance matrix of  $\hat{\boldsymbol{\beta}}_{MI}$  is given by

$$\hat{\mathbf{U}}_{MI} = \tilde{\mathbf{W}} + \left( \frac{M+1}{M} \right) \tilde{\mathbf{B}},$$

where

$$\tilde{\mathbf{W}} = \frac{1}{M} \sum_{m=1}^M \tilde{\mathbf{U}}_m \quad \text{and} \quad \tilde{\mathbf{B}} = \frac{1}{M-1} \sum_{m=1}^M (\tilde{\boldsymbol{\beta}}_m - \hat{\boldsymbol{\beta}}_{MI})(\tilde{\boldsymbol{\beta}}_m - \hat{\boldsymbol{\beta}}_{MI})'.$$

### 3.2 Weighted Generalized Estimating Equations

Robins *et al.* (1995) proposed a class of weighted estimating equations to allow for MAR mechanism. In binary longitudinal data, Chen & Zhou (2011) extended the method to accommodate arbitrary patterns of missing response and missing covariate. Their method was adapted here for longitudinal ordinal responses.

Define a weight matrix  $\boldsymbol{\Delta}_i = [\delta_{itt'}]_{T_i(J_i-1) \times T_i(J_i-1)}$ ,  $t = 1, \dots, T_i$ ,  $t' = 1, \dots, T_i$ , where  $\delta_{itt'} = \{I(R_{it} = 1, R_{it'} = 3) + I(R_{it} = 3, R_{it'} = 3)\} / \pi_{itt'}$  for  $t \neq t'$ ,  $\delta_{itt} = I(R_{it} = 3) / \pi_{it}$ , and  $\pi_{itt'} = Pr(R_{it} = 1, R_{it'} = 3 | \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i) + Pr(R_{it} = 3, R_{it'} = 3 | \mathbf{O}_i, \mathbf{X}_i, \mathbf{Z}_i)$ . Let  $\mathbf{M}_i = \mathbf{F}_i^{-1/2} (\mathbf{C}_i^{-1} \cdot \boldsymbol{\Delta}_i) \mathbf{F}_i^{-1/2}$  where  $\mathbf{A} \cdot \mathbf{B} = [a_{it} \cdot b_{it}]$  denotes the Hadamard product of matrix  $\mathbf{A} = [a_{it}]$  and  $\mathbf{B} = [b_{it}]$ .

The weighted generalized estimating equations (WGEE) for  $\boldsymbol{\beta}$  are given by

$$\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\psi}) = \sum_{i=1}^n \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\psi}) = \mathbf{0}, \quad (7)$$

where  $\mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\psi}) = \mathbf{D}_i \mathbf{M}_i (\mathbf{Y}_i - \boldsymbol{\mu}_i)$ . A consistent estimate for  $\boldsymbol{\beta}$  can be obtained by solving (7), under the correct specification of the missing data model.

To model  $\lambda_{itk}$  it is adopted a politomic logistic regression, with  $\lambda_{it0}$  as the reference category, that is

$$\log \left( \frac{\lambda_{itk}}{\lambda_{it0}} \right) = \mathbf{u}_{itk}^T \boldsymbol{\psi}_k, \quad k = 1, 2, 3, \quad (8)$$

where the covariates  $\mathbf{u}_{itk}$  are some function of  $\{\bar{\mathbf{R}}_{it}, \bar{\mathbf{O}}_{it}^o, \bar{\mathbf{X}}_{it}^o, \bar{\mathbf{Z}}_{it}\}$ .

## 4 Doubly Robust GEE for Longitudinal Ordinal Data

Some authors (e.g., Scharfstein *et al.* (1999), Tsiatis (2006)) noted that adding a term of expectation zero, say  $\phi(\cdot)$ , to the inverse probability weighted estimators would still

result in consistent estimates under a MAR mechanism. The solutions of these augmented estimating equations give rise to the so-called *doubly robust* estimators.

Chen & Zhou (2011) showed that the optimal  $\phi_{opt}$  for missing response and covariate is given by  $\phi_{opt} = E_{(\mathbf{Y}_i^m, \mathbf{X}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i, \mathbf{R}_i)} \{ \mathbf{D}_i \mathbf{N}_i (\mathbf{Y}_i - \boldsymbol{\mu}_i) \}$ , with  $\mathbf{N}_i = \mathbf{F}_i^{-1/2} \{ \mathbf{C}_i^{-1} \cdot (\mathbf{1}\mathbf{1}^T - \boldsymbol{\Delta}_i) \} \mathbf{F}_i^{-1/2}$ , where  $\mathbf{1}$  is a vector of 1's of length  $T_i(J-1)$ , and  $\mathbf{Y}_i^m$  and  $\mathbf{X}_i^m$  denote the missing components of  $\mathbf{Y}_i$  and  $\mathbf{X}_i$ , respectively.

An improved estimate for  $\boldsymbol{\beta}$  can then be obtained by solving the estimating equations

$$\mathbf{S}_1(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{S}_{1i}(\boldsymbol{\theta}) = \sum_{i=1}^n [\mathbf{D}_i \mathbf{M}_i (\mathbf{Y}_i - \boldsymbol{\mu}_i) + E_{(\mathbf{Y}_i^m, \mathbf{X}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i, \mathbf{R}_i)} \{ \mathbf{D}_i \mathbf{N}_i (\mathbf{Y}_i - \boldsymbol{\mu}_i) \}] = \mathbf{0}. \quad (9)$$

The estimator for  $\boldsymbol{\beta}$  in (9) is doubly-robust in the sense that it is consistent if *at least one* of the missing data model or the covariate model is correctly specified.

Applications of doubly robust estimators in longitudinal settings include Bang & Robins (2005), Seaman & Copas (2009), Chen & Zhou (2011) and Birhanu *et al.* (2011). Those developments focus mainly on binary response and have not been, to our knowledge, investigated with ordinal longitudinal data. In this work it is considered a longitudinal response measured on an ordinal scale.

The referred expectation in the second part of (9) is over the conditional distribution of  $(\mathbf{Y}_i^m, \mathbf{X}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i, \mathbf{R}_i)$ , which can be written as

$$\begin{aligned} P(\mathbf{Y}_i^m = \mathbf{y}_i^m, \mathbf{X}_i^m = \mathbf{x}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i, \mathbf{R}_i; \boldsymbol{\beta}^*, \boldsymbol{\gamma}) &= P(\mathbf{Y}_i^m = \mathbf{y}_i^m, \mathbf{X}_i^m = \mathbf{x}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i; \boldsymbol{\beta}^*, \boldsymbol{\gamma}) \\ &= P(\mathbf{Y}_i^m = \mathbf{y}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i = \mathbf{x}_i, \mathbf{Z}_i; \boldsymbol{\beta}^*) \\ &\quad \times P(\mathbf{X}_i^m = \mathbf{x}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i; \boldsymbol{\gamma}). \end{aligned}$$

The multivariate distribution  $P(\mathbf{Y}_i^m = \mathbf{y}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i = \mathbf{x}_i, \mathbf{Z}_i; \boldsymbol{\beta}^*)$  is expressed through a product of univariate ordinal models.

## 4.1 Estimation for the Nuisance Parameters

The method of maximum likelihood is employed to estimate  $\boldsymbol{\psi}$ . The log likelihood of the politomic logistic model for  $\boldsymbol{\psi}$  has the form

$$l(\boldsymbol{\psi}) = \sum_{i=1}^n l_i(\boldsymbol{\psi}) = \sum_{i=1}^n \sum_{t=1}^{T_i} \sum_{k=0}^3 I(R_{it} = k) \log(\lambda_{itk}),$$

with corresponding score function given by

$$\mathbf{S}_2(\boldsymbol{\psi}) = \sum_{i=1}^n \mathbf{S}_{2i}(\boldsymbol{\psi}) = \sum_{i=1}^n \sum_{t=1}^{T_i} \sum_{k=0}^3 \frac{I(R_{it} = k)}{\lambda_{itk}} \frac{\partial \lambda_{itk}}{\partial \boldsymbol{\psi}^T}.$$

The maximum likelihood estimator  $\hat{\boldsymbol{\psi}}$ , is obtained by solving  $\mathbf{S}_2(\boldsymbol{\psi}) = \mathbf{0}$ .

For the missing covariate model, the observed likelihood function for  $\boldsymbol{\gamma}$  is

$$L_3(\boldsymbol{\gamma}) = \prod_{i=1}^n \int Pr(\mathbf{X}_i | \mathbf{Z}_i, \mathbf{Y}_i) d\mathbf{X}_i^m,$$

with score function  $\mathbf{S}_3(\boldsymbol{\gamma}) = \sum_{i=1}^n \mathbf{S}_{3i}(\boldsymbol{\gamma})$ , where  $\mathbf{S}_{3i}(\boldsymbol{\gamma}) = \partial \log \int Pr(\mathbf{X}_i | \mathbf{Z}_i, \mathbf{Y}_i) d\mathbf{X}_i^m / \partial \boldsymbol{\gamma}^T$ . Similarly, a consistent estimator of  $\boldsymbol{\gamma}$  can be obtained by solving  $\mathbf{S}_3(\boldsymbol{\gamma}) = \mathbf{0}$ .

## 4.2 Estimation and Inference for the Doubly Robust Method

Let's denote the vector of all parameters as  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\psi}^T, \boldsymbol{\gamma}^T)^T$ . Our primary interest lies in estimating  $\boldsymbol{\beta}$ . Such task can be accomplished by plugging in the estimates  $\hat{\boldsymbol{\psi}}$  and  $\hat{\boldsymbol{\gamma}}$  in (9) and solving the estimating equations for  $\boldsymbol{\beta}$ , that is,

$$\mathbf{S}_1(\boldsymbol{\beta}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\gamma}}) = \sum_{i=1}^n \mathbf{S}_{1i}(\boldsymbol{\beta}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\gamma}}) = 0. \quad (10)$$

Second term of  $\mathbf{S}_{1i}$  can be written, for  $\mathbf{X}$  discrete, as

$$E_{(\mathbf{Y}_i^m, \mathbf{X}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i, \mathbf{R}_i)} \{ \mathbf{D}_i \mathbf{N}_i(\mathbf{Y}_i - \boldsymbol{\mu}_i) \} = \sum_{(\mathbf{y}_i^m, \mathbf{x}_i^m)} w_{ixy} \{ \mathbf{D}_i \mathbf{N}_i(\mathbf{Y}_i - \boldsymbol{\mu}_i) \},$$

where the weight  $w_{ixy}$  is given by

$$w_{ixy} = P(\mathbf{Y}_i^m = \mathbf{y}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i = \mathbf{x}_i, \mathbf{Z}_i; \boldsymbol{\beta}^*) \times P(\mathbf{X}_i^m = \mathbf{x}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i; \hat{\boldsymbol{\gamma}}).$$

In the case of  $\mathbf{X}$  continuous, the second term in  $\mathbf{S}_{1i}$  takes the form

$$E_{(\mathbf{Y}_i^m, \mathbf{X}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i, \mathbf{R}_i)} \{ \mathbf{D}_i \mathbf{N}_i(\mathbf{Y}_i - \boldsymbol{\mu}_i) \} = \int_{(\mathbf{Y}_i^m, \mathbf{X}_i^m)} w_{ixy} \{ \mathbf{D}_i \mathbf{N}_i(\mathbf{Y}_i - \boldsymbol{\mu}_i) \} d\mathbf{Y}_i^m d\mathbf{X}_i^m,$$

with conditional probability

$$w_{ixy} = P(\mathbf{Y}_i^m = \mathbf{y}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i = \mathbf{x}_i, \mathbf{Z}_i; \boldsymbol{\beta}^*) \times P(\mathbf{X}_i^m = \mathbf{x}_i^m | \mathbf{Y}_i^o, \mathbf{X}_i^o, \mathbf{Z}_i; \hat{\boldsymbol{\gamma}}).$$

The expectation in (9) can be cumbersome, depending on the missing data pattern. In such case, instead of using numerical integration, a Monte Carlo method can be applied to approximate the corresponding integral.

In this work it is assumed independence working correlation and it is adopted a sandwich standard error as given in Appendix.

## 5 Simulation Study

A small simulation study taking into account different sample sizes was conducted in order to quantify the bias and precision under misspecification of the predictive models. It is considered a study with  $T_i = T = 3$  repeated ordinal measures (with three categories) and two covariates (quantitative and qualitative). The true marginal model is

$$\text{logit } Pr(O_{it} \leq j | X_{it}, Z_{it}) = \beta_{0j} + \beta_1 X_{it} + \beta_2 Z_{it}, \quad j = 1, 2. \quad (11)$$

where  $Z_{it}$  is normal with unity variance and mean  $(0, 0.5, 1)$  for  $t = 1, 2, 3$ .

The binary covariate  $X_{it}$  may be missing at some time points and is generated according to

$$\text{logit } Pr(X_{it} = 1 | \bar{X}_{it}, Z_{it}) = \gamma_0 + \gamma_1 X_{i,t-1} + \gamma_2 Z_{it}. \quad (12)$$

It is assumed  $\beta_{01} = -0.4$ ,  $\beta_{02} = 1.2$ ,  $\beta_1 = -0.5$ ,  $\beta_2 = 0.5$ ,  $\gamma_0 = \log(1)$ ,  $\gamma_1 = 2$  and  $\gamma_2 = 2$ . The correlated ordinal responses were generated according to the algorithm proposed by



Touloumis (Anestis Touloumis, 2013) with constant correlation between the latent vectors set equal to  $\rho = 0.9$ .

As independent estimating equations were fitted,  $R_{it}$  can be defined as the indicator of observing both  $O_{it}$  and  $X_{it}$ , and it was taken

$$\log \left( \frac{Pr(R_{it} = 1)}{Pr(R_{it} = 0)} \right) = \psi_{0t} + \psi_1 I(R_{i,t-1} = 1) + \psi_2 O_{i,t-1}^* + \psi_3 X_{i,t-1}^* + \psi_4 Z_{it}, \quad t = 2, 3, \quad (13)$$

where  $O_{i,t-1}^* = O_{i,t-1}$ , if  $O_{i,t-1}$  is observed and 0 otherwise, and  $X_{i,t-1}^* = X_{i,t-1}$  if  $X_{i,t-1}$  is observed and 0 otherwise. The true values are taken as  $\psi_{02} = 6.6$ ,  $\psi_{03} = 6$ ,  $\psi_1 = 2$ ,  $\psi_2 = -2$ ,  $\psi_3 = -2$  and  $\psi_4 = 2$ . It was observed about 24% of missing observations under this setup.

For comparison purposes, it was considered ordinary GEE for the complete and available data, respectively, weighted GEE (WGEE), multiple imputation (MIGEE) by chained equations (van Buuren & Groothuis-Oudshoorn, 2011) with  $M = 10$ , and the proposed doubly robust version (DRGEE). In order to investigate robustness of these methods, the predicted models were also misspecified by omitting the covariate  $X_{t-1}$  from the covariate model (12) or the missing data model (13).

Results are summarized in Table 1. In each of the  $S = 1000$  Monte Carlo replications it was obtained the relative bias percentage for each parameter, defined as  $100 \times (\hat{\beta} - \beta) / \beta$ , its standard deviation obtained through the sandwich estimator, and the coverage probability as a nominal 95%.

Specially, MAR missingness impact over the response and the covariate is observed for all the regression parameters, the largest relative bias occur in binary covariate  $X$ . This comes in addition to the natural increase of parameter uncertainty. Bias in the intercept coefficients imply incorrect predicted probabilities for the levels of the response, whereas bias for parameter estimates associated with the regression covariates may erroneously attenuate or highlight an effect, thus leading to misinterpretations related to the importance of a given predictor on the longitudinal dynamics of the ordinal response.

It can be observed that for small sample size even the GEE with for complete data presents a certain degree of bias. Increasing the sample size allows to clarify the performance distinctions among the compared methods. WGEE and MIGEE methods are valid when the model for the weight or the imputation model, respectively, are correctly specified. In this case it is noted that both methods give good results for large sample sizes, the main distinction between them being due to the greater variability of the estimates for the weighted estimator.

Doubly robust method requires the simultaneous specification of two predictive model. When at least one of them is correctly specified the resulting estimator is still consistent. Estimates are, on average, closer to those obtained with fully observed data compared to WGEE or MIGEE. This behavior is systematic and it can be observed to all parameters. By increasing the sample size the estimates from DRGEE present empirical bias in general smaller than their single robust competitors. This is specially true for the parameter associated with the incomplete binary variable  $X$ . Regarding the uncertainty of parameter estimates, it is noted that the variability in DRGEE is greater than multiple imputation, but of the same order as of the weighted method. Further, the efficiency of the doubly robust estimates appears relatively more sensitive to misspecification of the weight model than the covariate model. Empirical coverage rates were acceptable for correctly specified

Table 1: Relative bias percentage, standard deviation and empirical coverage for 1000 simulations of incomplete covariate and response data.

	Empirical Bias				Standard Deviation				Empirical Coverage			
	$\beta_{01}$	$\beta_{02}$	$\beta_1$	$\beta_2$	$\beta_{01}$	$\beta_{02}$	$\beta_1$	$\beta_2$	$\beta_{01}$	$\beta_{02}$	$\beta_1$	$\beta_2$
$n = 50$												
Complete	-4.60	7.89	15.07	7.38	0.379	0.410	0.458	0.179	0.94	0.94	0.97	0.95
Available	-19.81	15.37	-8.36	-11.66	0.386	0.420	0.473	0.190	0.94	0.92	0.97	0.94
WGEE( $r^+$ )	-12.33	12.02	19.96	1.74	0.406	0.440	0.513	0.204	0.94	0.94	0.97	0.95
WGEE( $r^-$ )	1.50	7.43	1.31	0.51	0.413	0.450	0.508	0.201	0.95	0.93	0.96	0.95
MIGEE( $x^+$ )	22.17	-3.44	-38.26	-4.73	0.386	0.417	0.480	0.186	0.96	0.96	0.97	0.96
MIGEE( $x^-$ )	15.81	-0.13	-26.18	-2.54	0.387	0.417	0.481	0.187	0.95	0.96	0.97	0.96
DRGEE( $x^+, r^+$ )	-8.86	9.67	22.20	8.99	0.410	0.442	0.513	0.199	0.95	0.94	0.97	0.95
DRGEE( $x^-, r^+$ )	-7.77	9.44	19.33	7.91	0.471	0.507	0.579	0.204	0.95	0.94	0.97	0.95
DRGEE( $x^+, r^-$ )	-7.95	10.05	20.58	8.26	0.433	0.473	0.540	0.197	0.95	0.93	0.96	0.95
DRGEE( $x^-, r^-$ )	2.86	6.42	-0.28	2.99	0.414	0.450	0.529	0.200	0.95	0.93	0.97	0.95
$n = 150$												
Complete	-1.36	1.53	5.13	1.99	0.220	0.236	0.264	0.103	0.94	0.94	0.96	0.95
Available	-17.50	8.94	-18.41	-17.15	0.224	0.242	0.273	0.109	0.91	0.92	0.94	0.91
WGEE( $r^+$ )	-5.84	3.41	13.47	1.36	0.247	0.267	0.317	0.126	0.93	0.93	0.95	0.94
WGEE( $r^-$ )	10.20	-2.04	-12.91	-2.59	0.253	0.279	0.307	0.120	0.94	0.93	0.93	0.94
MIGEE( $x^+$ )	8.22	-2.57	-14.36	-2.40	0.228	0.244	0.283	0.108	0.94	0.94	0.96	0.95
MIGEE( $x^-$ )	9.08	-1.97	-16.33	-3.60	0.226	0.242	0.280	0.108	0.93	0.93	0.95	0.94
DRGEE( $x^+, r^+$ )	-4.21	2.64	11.90	4.12	0.249	0.269	0.317	0.117	0.94	0.94	0.96	0.95
DRGEE( $x^-, r^+$ )	-5.02	3.07	12.26	3.63	0.320	0.345	0.396	0.119	0.94	0.94	0.95	0.95
DRGEE( $x^+, r^-$ )	-1.86	1.99	7.26	2.95	0.249	0.275	0.313	0.113	0.94	0.94	0.95	0.94
DRGEE( $x^-, r^-$ )	9.88	-1.86	-15.60	-2.86	0.246	0.269	0.317	0.116	0.94	0.93	0.94	0.94
$n = 300$												
Complete	1.42	0.19	-0.04	0.85	0.156	0.167	0.187	0.072	0.96	0.94	0.94	0.95
Available	-14.82	7.67	-23.24	-18.14	0.159	0.171	0.194	0.077	0.93	0.91	0.92	0.87
WGEE( $r^+$ )	-0.24	0.75	5.61	1.23	0.182	0.198	0.235	0.095	0.95	0.94	0.94	0.93
WGEE( $r^-$ )	15.64	-4.50	-22.17	-3.89	0.184	0.205	0.222	0.088	0.95	0.94	0.92	0.94
MIGEE( $x^+$ )	6.23	-1.86	-9.95	-1.44	0.162	0.174	0.201	0.076	0.95	0.94	0.94	0.95
MIGEE( $x^-$ )	10.27	-2.70	-18.22	-3.89	0.160	0.172	0.200	0.076	0.94	0.92	0.92	0.94
DRGEE( $x^+, r^+$ )	-0.06	0.83	2.82	1.28	0.185	0.202	0.242	0.089	0.96	0.95	0.95	0.94
DRGEE( $x^-, r^+$ )	-0.71	1.08	3.84	1.43	0.194	0.213	0.251	0.088	0.96	0.95	0.95	0.94
DRGEE( $x^+, r^-$ )	2.53	0.03	-2.11	0.12	0.178	0.198	0.224	0.082	0.96	0.96	0.94	0.94
DRGEE( $x^-, r^-$ )	13.86	-3.69	-24.29	-5.51	0.176	0.194	0.228	0.084	0.95	0.94	0.93	0.94
$n = 600$												
Complete	-1.01	0.57	1.93	0.82	0.110	0.118	0.132	0.051	0.95	0.93	0.94	0.94
Available	-16.78	7.85	-22.20	-18.43	0.112	0.121	0.137	0.054	0.91	0.88	0.92	0.78
WGEE( $r^+$ )	-1.00	0.28	5.43	1.25	0.132	0.144	0.170	0.070	0.95	0.94	0.95	0.95
WGEE( $r^-$ )	14.87	-4.99	-21.53	-3.03	0.133	0.149	0.158	0.064	0.94	0.94	0.92	0.93
MIGEE( $x^+$ )	1.76	-0.66	-3.63	-0.39	0.115	0.123	0.143	0.054	0.95	0.94	0.95	0.94
MIGEE( $x^-$ )	6.96	-2.01	-14.18	-3.38	0.113	0.122	0.141	0.054	0.94	0.93	0.93	0.93
DRGEE( $x^+, r^+$ )	-1.49	0.59	3.19	0.98	0.132	0.144	0.174	0.062	0.96	0.94	0.94	0.94
DRGEE( $x^-, r^+$ )	-1.39	0.56	2.96	0.91	0.136	0.149	0.178	0.062	0.96	0.94	0.96	0.95
DRGEE( $x^+, r^-$ )	0.42	0.08	-0.12	0.47	0.126	0.140	0.159	0.058	0.96	0.95	0.95	0.93
DRGEE( $x^-, r^-$ )	12.24	-3.82	-23.18	-5.35	0.124	0.138	0.161	0.060	0.93	0.93	0.92	0.93

“+” indicates correctly specified model and “-” indicates misspecified model omitting the  $X_t$  predictor

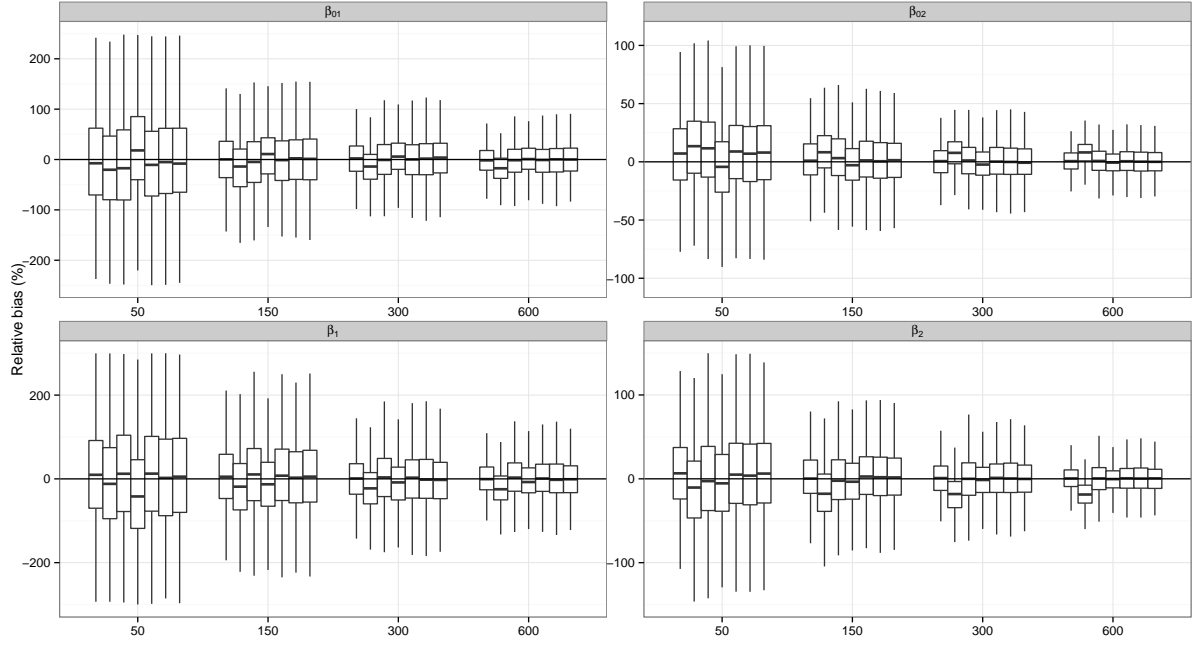


Figure 1: Boxplot of the relative bias for parameter estimates under correctly specified models

WGEE and MIGEE as well as for DRGEE when at least one of predictive models are correctly specified.

Figure 1 shows boxplots of the percentage relative bias for the methods expected to be valid. The boxes represent, respectively, GEE for complete and available data, WGEE( $r^+$ ), MIGEE( $x^+$ ), DRGEE( $x^+, r^+$ ), DRGEE( $x^-, r^+$ ) and DRGEE( $x^+, r^-$ ). As the degree of bias is different in the parameter estimates, for easy of visualization, they are represented in different scales. It can be seen that the estimates with larger variability were those associated with the first intercept and the covariate with incomplete values. Proposed method presents median relative bias close to zero and very similar to those observed with complete data for all sample sizes. Variability of the DR estimators is slightly larger than MI but of the same order as the weighted estimator. As expected, for all methods it can be noticed a decrease in the relative bias with increasing sample size, reflecting their theoretical asymptotic consistency.

Figure 2 allows the comparison of methods incorrectly specified. The boxes represent, respectively, GEE for complete and available data, WGEE( $r^-$ ), MIGEE( $x^-$ ), and DRGEE( $x^-, r^-$ ).

When a key covariate is omitted from the weight model and/or the imputation one, it is expected all methods to be biased. This is specially true for the first intercept and for the incomplete binary covariate. Bias for MIGEE( $x^-$ ) seemed to get smaller than WGEE( $r^-$ ) as sample sizes increases. Bias for DRGEE( $x^-, r^-$ ) was comparable to WGEE( $r^-$ ) and slightly higher than multiple imputation for large sample sizes. In terms of variability of the estimates, the same pattern is observed as with the correctly specified models. That is, the multiple imputation is more efficient, followed by the doubly robust estimator and the weighed estimator.

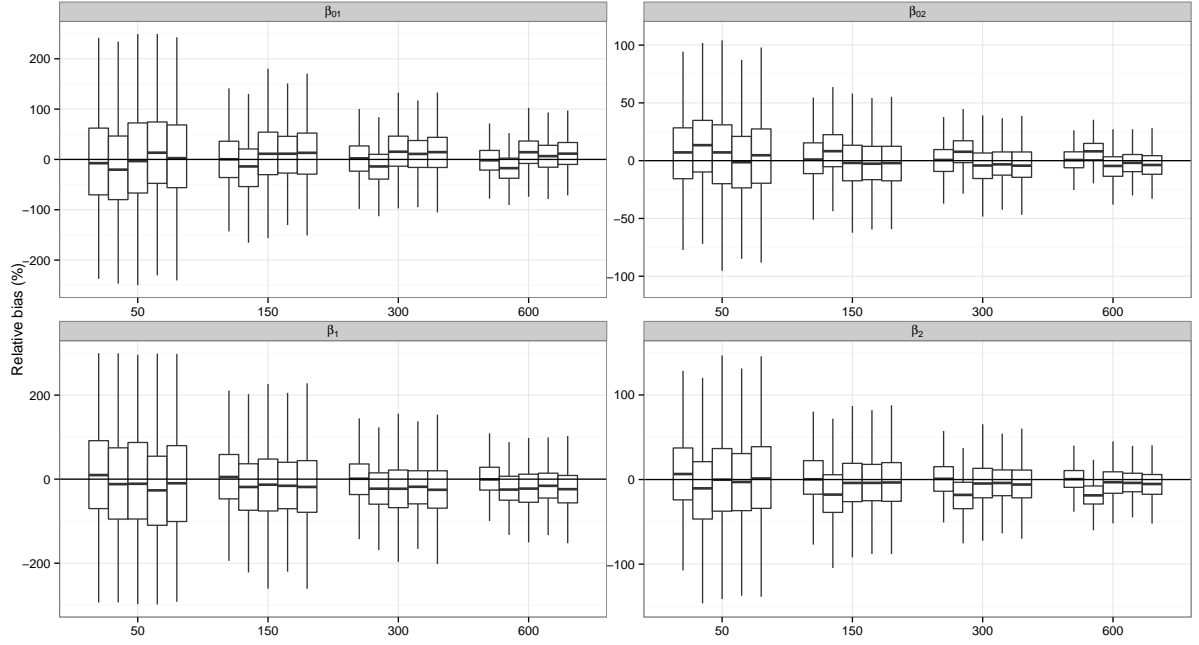


Figure 2: Boxplot of the relative bias for the estimates with incorrectly specified models

## 6 Data Analysis: Analgesia in Childbirth

This study was conducted in Minas Gerais state, Brazil, in order to compare two techniques of analgesia for labor pain. There were 49 patients who were monitored during their entire labor period until childbirth. Pain intensity was subjectively assessed by the patient, and measurements of blood pressure, maternal heart rate, consumption of oxytocin, sedation level, signs of respiratory depression, apnea, and other variables were recorded. One of the techniques used was epidural analgesia (the gold standard), which is a local anesthetic. The other one, whose efficiency was to be compared to the gold standard, involved continuous intravenous infusion of remifentanyl.

The response of interest is the intensity of pain as measured by a Visual Analog Scale (VAS) (1: tolerable and mild pain; 2: moderate pain that causes discomfort; 3: intense and unbearable pain). Three measurements (0, 60, 90 minutes) were selected for data analysis. Predictor variables considered were treatment GROUP (0: peridural; 1: remifentanyl), AGE (in years), DU (uterine dilatation), and OXYT (consumption of oxytocin). The OXYT is a time-varying ordinal covariate, coded as 1, if no consumption, 2, if consumption equals to 10 or 30, and 3, if consumption equals or above 45. These two covariates were chosen after a previous exploratory analysis.

The response and oxytocin consumption was missing for 9 patients at the time 60 and for 18 patients at time 90. Missing is due to childbirth happened before 60 or 90 minutes. Therefore a MAR mechanism seems to be a reasonable assumption for this data set. The other covariates in the analysis were fully observed.

For the ordinal response it was used the following proportional odds model

$$\text{logit } Pr(PAIN_{itj} \leq j | \mathbf{u}_{it}) = \beta_{0j} + \mathbf{u}_{it}^T \boldsymbol{\beta}, \quad j = 1, 2, \quad t = 1, 2, 3, \quad (14)$$

where  $\mathbf{u}_{it}$  is the covariate vector at time  $t$ , and it is formed by TIME, GROUP, AGE,

Table 2: Regression Parameters for the Analgesia in Birth Data

Parameter	Available			WGEE			MIGEE			DRGEE		
	Est	SE	P	Est	SE	P	Est	SE	P	Est	SE	P
INTERCEPT1	1.796	1.308	0.170	1.827	1.267	0.149	1.599	1.237	0.196	1.649	1.315	0.210
INTERCEPT2	3.302	1.314	0.012	3.261	1.310	0.013	3.105	1.242	0.012	3.069	1.351	0.023
TIME	-0.182	0.286	0.524	-0.157	0.315	0.619	-0.110	0.273	0.687	-0.192	0.293	0.511
GROUP	-1.221	0.445	0.006	-1.244	0.439	0.005	-1.056	0.409	0.010	-1.008	0.390	0.010
AGE	-0.066	0.035	0.057	-0.072	0.033	0.027	-0.062	0.031	0.046	-0.071	0.035	0.045
DU	-0.362	0.158	0.022	-0.354	0.168	0.035	-0.372	0.156	0.017	-0.340	0.166	0.040
OXYT(=2)	0.727	0.554	0.189	0.712	0.554	0.199	0.954	0.542	0.078	0.821	0.531	0.122
OXYT(=3)	1.285	0.510	0.012	1.431	0.503	0.004	1.410	0.488	0.004	1.468	0.475	0.002

DU and OXIT. In response model, TIME predictor was expressed in hours rather than minutes.

When using WGEE or DRGEE it is necessary to correctly model  $\pi_i$  in order to obtain consistent estimates of  $\beta$ . For the missing data process  $R_{it}$  was defined as the indicator of observing both  $PAIN_{it}$  and  $OXIT_{it}$ , and take the following form

$$\log \left( \frac{Pr(R_{it} = 1)}{Pr(R_{it} = 0)} \right) = \psi_{0t} + \mathbf{w}_{it}^T \boldsymbol{\psi}, \quad t = 2, 3, \quad (15)$$

where  $\mathbf{w}_{it}$  includes GROUP, AGE, DU, histories of OXYT and PAIN, and the previous indicator of missing data.

The distribution of the missing covariate OXYT also needs to be specified in a predictive model. With this aim, it was assumed a proportional odds model of the form

$$\text{logito } Pr(OXYT_{itj} \leq j | \mathbf{v}_{it}) = \gamma_{0j} + \mathbf{v}_{it}^T \boldsymbol{\gamma}, \quad j = 1, 2, \quad t = 2, 3 \quad (16)$$

where  $\mathbf{v}_{it}$  includes main effects for GROUP, AGE, and DU. It can be observed that the estimate of  $\boldsymbol{\gamma}$  is not of interest, however it is necessary to model the missing mechanism related to covariate as close as possible to true in order to obtain valid estimates of  $\beta$ . The same is true for the missing data process. All predictors in this model process were maintained since an overspecification is better than a underspecification.

Results from four methods are shown in Table 2. The first one is the usual GEE method using the available data; the second is the weighted method (WGEE) using model (15) for the weights; the third is the multiple imputation by chained equation (MIGEE) in the *R* package *mice*; and the fourth, labeled DRGEE, is the proposed doubly robust method using (15) and (16) for the weight and the covariate models, respectively. It was used an independent working correlation.

TIME effect is non significant for all the four methods. All methods provide the same conclusion for effects of GROUP. The negative effect for GROUP means that the chance of women feel mild pain is lower among the group receiving the remifentanyl compared to the peridural group (the estimated odds is  $e^{-1.008} = 0.365$  in the doubly robust method). All methods also agree with respect to the effect of DU. That is, for each increase of 1 cm of uterine dilation the chance of the parturient feel mild pain decreases (in the DRGEE it is  $e^{-0.340} = 0.712$ , for example). It can be noticed that  $p$ -value for AGE effect goes from a non-significant of 0.057 in the standard GEE to a significant one in DRGEE, as well as for the other two missing data approaches. The conclusion is that older women have lower chance of experiencing mild pain than young women. For the OXIT covariate all methods reached the same conclusion.

## 7 Discussion

When longitudinal ordinal data are of interest, doubly robust estimator is a nice alternative. Doubly robust method combines ideas from weighting and imputation and has been applied elsewhere for estimation of means, causal inference, and in the longitudinal setting for binary response data (Bang & Robins (2005), Carpenter *et al.* (2006), Seaman & Copas (2009), Chen & Zhou (2011), Li *et al.* (2013)). However, as far as we know, it has not been investigated for the longitudinal ordinal case. A doubly robust estimator is attractive in the sense that it needs only the correct specification of at least one of the models, but not necessarily both. Simulation results have indicated that, when at least the covariate model or missing data model is correct, the doubly robust estimators are consistent and present small-sample bias comparable to single robust alternatives MIGEE or WGEE. The proposed method presented good coverage probabilities, as well as its competitors but with a slight larger variance than multiple imputation. Simulation results also indicated that the bias of doubly robust estimators when both the covariate model and the missing data model are incorrect was the same magnitude of misspecified WGEE or MIGEE. We hope that, in practical applications, none of the predictive models would be grossly misspecified and then the proposed estimator would have a great potential of reducing the bias if the MAR assumption is correct.

When the assumed independent working correlation structure differs from the true underlying structure, there is no price to pay in terms of the consistency and asymptotic normality of  $\beta$ , but such a poor choice may result in loss of efficiency (Molenberghs & Verbeke, 2005). However, modeling of the association structure in the presence of missing data remains a challenge, specially with longitudinal ordinal data, because there is no direct way of modeling the association parameters. Future research involves the investigation of the impact of other association structures in the doubly robust estimates.

In the proposed doubly robust estimator marginal means were modeled by cumulative logits. This implies a proportional odds model that in some cases may not be valid. Another future possible extension of the proposed model is, therefore, to allow non-proportional odds for a subset of the explanatory variables (Peterson & Harrell Jr, 1990).

## References

- Agresti, Alan. 2013. *Categorical data analysis*. John Wiley & Sons.
- Anestis Touloumis. 2013. *SimCorMultRes: Simulating Correlated Multinomial Responses*. Version 1.1.
- Bang, Heejung, & Robins, James M. 2005. Doubly robust estimation in missing data and causal inference models. *Biometrics*, **61**(4), 962–973.
- Birhanu, Teshome, Molenberghs, Geert, Sotto, Cristina, & Kenward, Michael G. 2011. Doubly robust and multiple-imputation-based generalized estimating equations. *Journal of biopharmaceutical statistics*, **21**(2), 202–225.
- Carpenter, James R., & Kenward, Michael G. 2013. *Multiple Imputation and its Application*. Wiley & Sons.

- Carpenter, James R, Kenward, Michael G, & Vansteelandt, Stijn. 2006. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **169**(3), 571–584.
- Chen, Baojiang, & Zhou, Xiao-Hua. 2011. Doubly robust estimates for binary longitudinal data analysis with missing response and missing covariates. *Biometrics*, **67**(3), 830–842.
- Donneau, AF, Mauer, M, Molenberghs, Geert, & Albert, A. 2014a. A simulation study comparing multiple imputation methods for incomplete longitudinal ordinal data. *Communications in Statistics-Simulation and Computation*.
- Donneau, Anne-Françoise, Mauer, Murielle, Lambert, Philippe, Molenberghs, Geert, & Albert, Adelin. 2014b. Simulation-based study comparing multiple imputation methods for non-monotone missing ordinal data in longitudinal settings. *Journal of Biopharmaceutical Statistics*.
- Li, Lingling, Shen, Changyu, Li, Xiaochun, & Robins, James M. 2013. On weighting approaches for missing data. *Statistical Methods in Medical Research*, **22**(1), 14–30.
- Liang, Kung-Yee, & Zeger, Scott L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, **73**(1), 13–22.
- Lipsitz, Stuart R, Kim, Kyungmann, & Zhao, Lueping. 1994. Analysis of repeated categorical data using generalized estimating equations. *Statistics in medicine*, **13**(11), 1149–1163.
- Little, Roderick JA, & Rubin, Donald B. 1987. *Statistical analysis with missing data*. Vol. 539. Wiley New York.
- Little, Roderick JA, & Rubin, Donald B. 2002. *Statistical analysis with missing data*.
- McCullagh, Peter. 1980. Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, 109–142.
- Molenberghs, Geert, & Verbeke, Geert. 2005. *Models for discrete longitudinal data*. Springer.
- Peterson, Bercedis, & Harrell Jr, Frank E. 1990. Partial proportional odds models for ordinal response variables. *Applied Statistics*, 205–217.
- Pierce, Donald A. 1982. The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statistics*, 475–478.
- Robins, James M, Rotnitzky, Andrea, & Zhao, Lue Ping. 1995. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**(429), 106–121.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. John Wiley.
- Schafer, Joseph L. 1997. *Analysis of incomplete multivariate data*. CRC press.

- Scharfstein, Daniel O, Rotnitzky, Andrea, & Robins, James M. 1999. Adjusting for non-ignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, **94**(448), 1096–1120.
- Seaman, Shaun, & Copas, Andrew. 2009. Doubly robust generalized estimating equations for longitudinal data. *Statistics in medicine*, **28**(6), 937–955.
- Toledano, Alicia Y, & Gatsonis, Constantine. 1999. Generalized estimating equations for ordinal categorical data: arbitrary patterns of missing responses and missingness in a key covariate. *Biometrics*, **55**(2), 488–496.
- Touloumis, Anestis, Agresti, Alan, & Kateri, Maria. 2013. GEE for Multinomial Responses Using a Local Odds Ratios Parameterization. *Biometrics*, 1–8.
- Tsiatis, Anastasios A. 2006. *Semiparametric theory and missing data*. Springer.
- van Buuren, S., Boshuizen, H. C., & Knook, D. L. 1999. Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis. *Statistics in Medicine*, **18**, 681–694.
- van Buuren, Stef. 2007. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, **16**(3), 219–242.
- van Buuren, Stef, & Groothuis-Oudshoorn, Karin. 2011. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, **45**(3), 1–67.



# A Appendix

## A.1 Asymptotic Variance

To state the asymptotic properties of  $\hat{\beta}$ , let  $\mathbf{S}_{1i}(\beta, \psi, \gamma)$  be the individual's contribution to the estimating equations for  $\beta$ ,  $\mathbf{S}_{2i}(\psi)$  be the individual's contribution to the estimating equations for  $\psi$ , and  $\mathbf{S}_{3i}(\gamma)$  be the individual's contribution to the estimating equations for  $\gamma$ .

Define  $\mathbf{\Gamma}(\beta, \psi, \gamma) = E \{ \partial \mathbf{S}_{1i}(\beta, \psi, \gamma) / \partial \beta^T \}$ ,  $\mathbf{I}_{12}(\beta, \psi, \gamma) = E \{ \partial \mathbf{S}_{1i}(\beta, \psi, \gamma) / \partial \psi^T \}$ ,  $\mathbf{I}_{13}(\beta, \psi, \gamma) = E \{ \partial \mathbf{S}_{1i}(\beta, \psi, \gamma) / \partial \gamma^T \}$ ,  $\mathbf{I}_2(\psi) = E \{ \partial \mathbf{S}_{2i}(\psi) / \partial \psi^T \}$ ,  $\mathbf{I}_3(\gamma) = E \{ \partial \mathbf{S}_{3i}(\gamma) / \partial \gamma^T \}$ , and  $\mathbf{Q}_i(\beta, \psi, \gamma) = \mathbf{S}_{1i}(\beta, \psi, \gamma) - \mathbf{I}_{12}(\beta, \psi, \gamma) \mathbf{I}_2^{-1}(\psi) \mathbf{S}_{2i}(\psi) - \mathbf{I}_{13}(\beta, \psi, \gamma) \mathbf{I}_3^{-1}(\gamma) \mathbf{S}_{3i}(\gamma)$ .

**Theorem 1** *If either the missing data model or the covariate model is correctly specified, then*

$$n^{1/2}(\hat{\beta} - \beta_0) \longrightarrow N(\mathbf{0}, \mathbf{\Gamma}^{-1}(\beta_0, \psi_0, \gamma_0) \mathbf{\Sigma} \{ \mathbf{\Gamma}^{-1}(\beta_0, \psi_0, \gamma_0) \}^T), \quad (17)$$

where  $\beta_0$  is the true value of  $\beta$ ,  $\psi_0$  and  $\gamma_0$  are the probability limits of  $\hat{\psi}$  and  $\hat{\gamma}$ , and  $\mathbf{\Sigma} = E \{ \mathbf{Q}_i(\beta_0, \psi_0, \gamma_0) \mathbf{Q}_i^T(\beta_0, \psi_0, \gamma_0) \}$ .

Inferences for  $\beta$  follows by replacing the unknown quantities in (17) by its consistent estimators. We make use of “generalized information equality” (Pierce, 1982) that

$E \{ \partial \mathbf{S}_{1i}(\beta, \psi, \gamma) / \partial \psi^T \} = -E \{ \mathbf{S}_{1i}(\beta, \psi, \gamma) \mathbf{S}_{2i}^T(\psi) \}$ , and  $E \{ \partial \mathbf{S}_{1i}(\beta, \psi, \gamma) / \partial \gamma^T \} = -E \{ \mathbf{S}_{1i}(\beta, \psi, \gamma) \mathbf{S}_{3i}^T(\gamma) \}$ . Similarly (Robins *et al.*, 1995),  $E \{ \partial \mathbf{S}_{2i}(\psi) / \partial \psi^T \} = -Var \{ \mathbf{S}_{2i}(\psi) \}$ , and  $E \{ \partial \mathbf{S}_{3i}(\gamma) / \partial \gamma^T \} = -Var \{ \mathbf{S}_{3i}(\gamma) \}$ .

The matrix  $\mathbf{\Gamma}$  is replaced by  $\hat{\mathbf{\Gamma}} = n^{-1} \sum_{i=1}^n \{ \partial \mathbf{S}_{1i}(\hat{\theta}) / \partial \beta^T \}$ , and  $\mathbf{\Sigma}$  by  $\hat{\mathbf{\Sigma}} = n^{-1} \sum_{i=1}^n \{ \hat{\mathbf{Q}}_i \hat{\mathbf{Q}}_i^T \}$ ,  $\hat{\mathbf{Q}}_i = \mathbf{S}_{1i}(\hat{\theta}) - \hat{\mathbf{I}}_{12}(\hat{\theta}) \hat{\mathbf{I}}_2^{-1}(\hat{\psi}) \mathbf{S}_{2i}(\hat{\psi}) - \hat{\mathbf{I}}_{13}(\hat{\theta}) \hat{\mathbf{I}}_3^{-1}(\hat{\gamma}) \mathbf{S}_{3i}(\hat{\gamma})$ ,  $\hat{\mathbf{I}}_{12}(\hat{\theta}) = n^{-1} \sum_{i=1}^n \{ \partial \mathbf{S}_{1i}(\hat{\theta}) / \partial \psi^T \}$ ,  $\hat{\mathbf{I}}_{13}(\hat{\theta}) = n^{-1} \sum_{i=1}^n \{ \partial \mathbf{S}_{1i}(\hat{\theta}) / \partial \gamma^T \}$ ,  $\hat{\mathbf{I}}_2(\hat{\psi}) = n^{-1} \sum_{i=1}^n \{ \partial \mathbf{S}_{2i}(\hat{\psi}) / \partial \psi^T \}$ ,  $\hat{\mathbf{I}}_3(\hat{\gamma}) = n^{-1} \sum_{i=1}^n \{ \partial \mathbf{S}_{3i}(\hat{\gamma}) / \partial \gamma^T \}$ .

The proof is similar to Chen & Zhou (2011) and is omitted here.